QA
76
M94
SI-Ref

M U D P I E no. 9

Museum and University Data, Program and Information Exchange

COMPUTER USE IN COMPILATION OF SCIENTIFIC REFERENCES

F. C. Kull, Burroughs Wellcome and Company, Tuckahoe, New York, has submitted a summary volume of the "Scientific papers published from 1930 to 1968 by the Wellcome Research Laboratories," 1969, 292 pp., which demonstrates the use of an IBM 360-40 in production of a listing of papers followed by an alphabetical index of authors, a "Keyword" index based on the titles, and finally a list of words in titles ignored by the computer in compiling the keyword index. Persons interested in the technique and the output are invited to contact Dr. Kull.

COMPUTER METHODS IN PLANT SYSTEMATICS

A two-year project at Michigan State titled "Computer Methods in Plant Systematics for Classroom Instruction and Associated Research" has been funded by NSF. The project is directed by John H. Beaman, with Stanwyn G. Shetler, Smithsonian Institution, as Cooperator and Larry E. Morse, M.S.U., a principal student participant. Lawrence T. Alexander, M.S.U., will serve as a systems development expert in the design and analysis of instructional systems.

Curricular development

To our knowledge there are at present no formal university courses in which the computer is used as an aid in the identification of a plant or animal. Our system, utilizing computerized matrices of taxa and characters, offers free choice in character selection and simultaneous use of any or all characters in the identification process. This computerized identification procedure also has the capability for suggesting character possibilities to the user that he might not have otherwise considered.

On-line computers make possible a dialogue between user and machine, which is essential to the most effective application of this and probably any other computerized identification system. Identification routines can be programmed to allow the computer to lead the student in making the identification and correct him when he makes a mistake. Used in this manner the student and the system can engage in an interaction which stimulates the student's interest in the identification process.

Each spring quarter Michigan State University offers a junior-level course (Botany 318) in plant identification and systematics. In 1969 this course had an enrollment of about 250 students. The lectures consider nomenclature, evolution, and characteristics of important families, while the laborabories are concerned primarily with the identification of trees and shrubs in winter condition (using Harlow's Twig Key) and the spring flora of Michigan (using Fassett's Spring Flora of Wisconsin). Computerized identification matrices are being prepared to be compatible with the texts, enabling a ready comparison between the new and tradi-

SMITHSONIAN
MAR 17 1970
LIBRARIES

SMITHSONIAN INSTITUTION MAR 10 1970

tional methods of identification. The laboratories will be provided with four teletype terminals, with each student allowed about one hour of terminal time. The students are supplied with a lecture syllabus and laboratory manual which includes instructions for use of the computer.

In alternate years the University offers a three-term course sequence in plant systematics for graduate students. In the first two courses of the sequence the evolution and classification of the angiosperms is emphasized and in the final course the philosophy of systematics is the principal subject. All three courses offer an excellent opportunity for introducing graduate students to computer methods in plant identification and the organization of taxonomic data. These students are now constructing data matrices for identification of plant groups in which they have a research interest. They may also become involved in computer program development for plant identification and taxonomic data handling, and can also employ batch processing methods for key construction and editing. The advanced courses also provide an opportunity for graduate students to obtain proficiency in computer identification techniques before they serve as teaching assistants in the introductory course. If interest proves sufficient, which seems likely, an advanced course concerned solely with computer methods in botany will be instituted.

Advanced Training and Research

As major participants in the project several graduate students will receive training in computer applications through assisting in the preparation of the computer-adapted instructional materials to be used in the courses. They are expected to use computer techniques in their research and to continue development of the prototype systems for key construction, key editing and printing, and on-line identification. The assembling of data to build a library of taxonomic data matrices is one of their prime responsibilities. Development of these software systems offer many challenging subjects for thesis research.

Inter-institutional sharing of resources and taxonomic data

One of the most important long-range implications of the project is the possibility for development of systems which will permit on-line sharing of taxonomic data among institutions. Such systems should make possible instantaneous and continual updating of taxonomic research data for use by all investigators.

The project participants will spend part of their time, especially in the summer, at the Smithsonian Institution where they will have the opportunity to utilize the collections of the U. S. National Herbarium. They will also profit from interaction with the scientific staff of the Smithsonian, which in turn should benefit by being able to draw upon the personnel resources of the University and by the increased use of its collections for research. Finally, the on-line link with the Smithsonian should increase students' appreciation of the dynamic importance of museum collections for systematic research.

Materials Available

A program package, not yet in final form, is available at cost for experimental use. Informal program descriptions, instructions for data preparation, and sample data and output are also available. [See below for list of programs.]

The participants will welcome comments and suggestions from the biological community.

John H. Beaman
Department of Botany and Plant Pathology
Michigan State University
East Lansing, Michigan 48823

20 January 1970

PROGRAM PACKAGE FOR COMPUTER AIDS TO IDENTIFICATION

A program package being developed by Larry E. Morse investigates the use of time-sharing computers as aids to identification of biological specimens. The techniques center on taxon/character data matrices, using the same data for all programs. Although final checking continues, the current versions of the various programs are available for experimental use, from Larry E. Morse, Department of Botany, Michigan State University, East Lansing, Michigan 48823, either as xeroxed lists or punched tape, at cost. Informal program descriptions, instructions for data preparation, and sample data and output are also available. A comprehensive article on the entire project should be completed soon, and a finalized "package" will be distributed then. The seven programs in the current set are described below. At present, each handles up to 20 character couplets for up to 50 taxa. All programs are in Fortran.

IDENT -- Identifies specimens on-line by conversational polyclave algorithm.

KEY -- Constructs and prints indented dichotomous identification keys from matrix data, by algorithm described in Amer. Jour. Botany 55:737 (1968).

DSCRBE-- Prints diagnostic descriptions of taxa, listing characters in order of decreasing importance.

CNTRST-- Contrasts two or more taxa, listing differentiating and similar characters.

QUIZ -- Queries students on most important characters of a taxon.

CHANGE-- Editing program for updating or correcting information in a stored data matrix.

TEST -- Tests the effectiveness of a matrix for identification and keying.

Several subroutines are shared by the various programs, including routines for reading a matrix from disk to core, selecting the most useful characters for differentiating a group of taxa, and ranking the characters of a particular taxon in order of importance.

Larry E. Morse
December 13, 1969

RECENT LITERATURE

Manning, Ray. A computer generated catalog of types: A byproduct of data processing. Curator, 12, 1969, 134-138.

Smith, J. M. Mathematical ideas in biology. Cambridge, 1968, vii + 152. Formulae for theoretical models in scale; population regulation; genetics; target theory; regulation and control; and diffusion.

Cheetham, A. H. and J. E. Hazel. Binary (Presence-absence) similarity coefficients. Journal of Paleontology, 43, 1969, 1130-36. A summary of all numerical coefficients, and a most useful table of comparisons of them.

New Journal: Computer Programs in Biomedicine. Editor: W. Schneider, Uppsala University Data Center, P. O. Box 2103, 5750 02, Uppsala 2, Sweden. Published by North Amsterdam Publishing Co., P. O. Box 3489, Amsterdam. Subscription: $30 per year.

De Tar, D. F., Editor. Computer programs for chemistry. Vol. 1. Benjamin, New York, 1968, xix + 208. A list of seven Fortran programs, highly specialized for chemistry (Transition energies of chemical shifts, first order kinetics, energy levels), but one is a plotting subroutine that may be of use to some biologists.

Field, J. G. The use of the information statistic in the numerical classification of heterogeneous systems. Journal of Ecology, 57, 1969, 565-569.

Pearcey, T. Computing systems for handling large-scale data. Australian Journal of Science, 31, 1969, 346-356.

Amacher, Peter et al. Preventing obsolescence of scientific reviews: An updated-review project. Science, 165, 1969, 1029-30. .Computer storage of non-published review material, available for criticism before final publication through rapid printing.

Gabrieli, E. R. [Consulting Editor.] The use of data mechanization and computers in clinical medicine. Annals of the New York Academy of Sciences, Vol. 161, Art. 2, Sept. 30, 1969, pp. 371-830. Don't let this title fool you - there are several papers here of value to biologists more broadly based than clinicians. For example: "Computer-produced distribution maps of disease," by H. C. Hopps, pp. 779-799; "Statistical techniques of difference," by J. M. Weiner and J. Marmorston, pp. 641-668; "Survey of pattern recognition," by W. J. Nilsson, pp. 380-401.

Rohlf, F. J. Grafpac, graphic output for the GE635 computer. Computer contribution 36, State Geological Survey, University of Kansas, Lawrence, 1969, pp. 1-49.

NETWORK INFORMATION SERVICE

General Electric has just announced a nationwide grid of communications and interconnected computer systems. A central computer in Cleveland can be accessed from any of 40 metropolitan areas, coast to coast. Material stored by one user can be accessed by a second user through previous agreement, and therefore the the first step in a museum network is already available to MUDPIE. Any institution contemplating installation of time-shared computer should look carefully at what GE offers, because this system or something resembling it must eventually form the basis for an information exchange. The Flora North America Project of the AIBS and Michigan State University are already engaged in planning of initial implementation of mutual access of each other's stored data (See note above on MSU's NSF grant).

FLORA NORTH AMERICA STAFF

The appointment of Harriet R. Meadow to the new position of Systems Development Manager of the Flora North America Project has just been announced by Dr. John R. Olive, Director of the American Institute of Biological Sciences.

A major objective of the Flora North America Project is to develop an information system to support floristic-taxonomic research. As the Systems Development Manager, Mrs. Meadow will be the "chief architect" of this system, and, in concert with the botanists of the project, will develop a system design and program implementation plan for the FNA.

Division of Reptiles and Amphibians
National Museum of Natural History
Smithsonian Institution
Washington, D. C. 20560